# Modelling for Science and Engineering
# Data Visualitzation and Modelling

## BAYESIAN NETWORKS

### Excercises Block 2

1. In Example 2 (see Figure 1), assume that $W$ is instantiated by $w_1$.

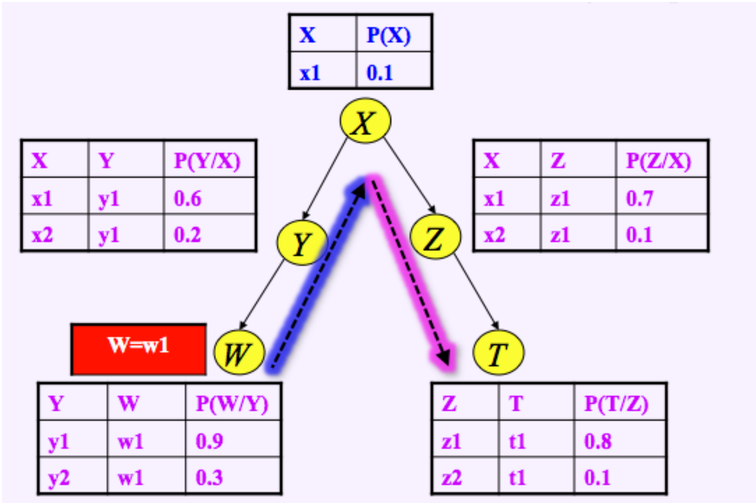


Figura 1: Example 2, Block 2

   (a) First compute $P(y_1/w_1)$ and $P(x_1/w_1)$ by using the **upward** propagation algorithm.

   (b) And then compute $P(z_1/w_1)$ and $P(t_1/w_1)$ by using the **downward** propagation algorithm.

2. Compute $P(x_1/t_2, w_1)$ assuming the Bayesian network of Example 2 (see Figure 1).

3. Repeat exercises 1 and 2 by using **gRain**, and check the results.

4. Using **gRain**, develop the Bayesian Network in Figure 2, used for the problem of detecting credit card fraud. The following variables are relevant for the problem:

   **Fraud (F):** whether the current purchase is fraudulent.

   **Gas (G):** whether gas has been purchased in the last 24 hours.

   **Jewelry (J):** whether jewelry has been purchased in the last 24 hours.

   **Age (A):** age of the card holder.

   **Sex (S):** sex of the card holder.

   These variables are all causally related. That is, a credit card thief is likely to buy gas and jewelry, and middle-aged women are most likely to buy jewelry, whereas young men are least likely to buy jewelry.

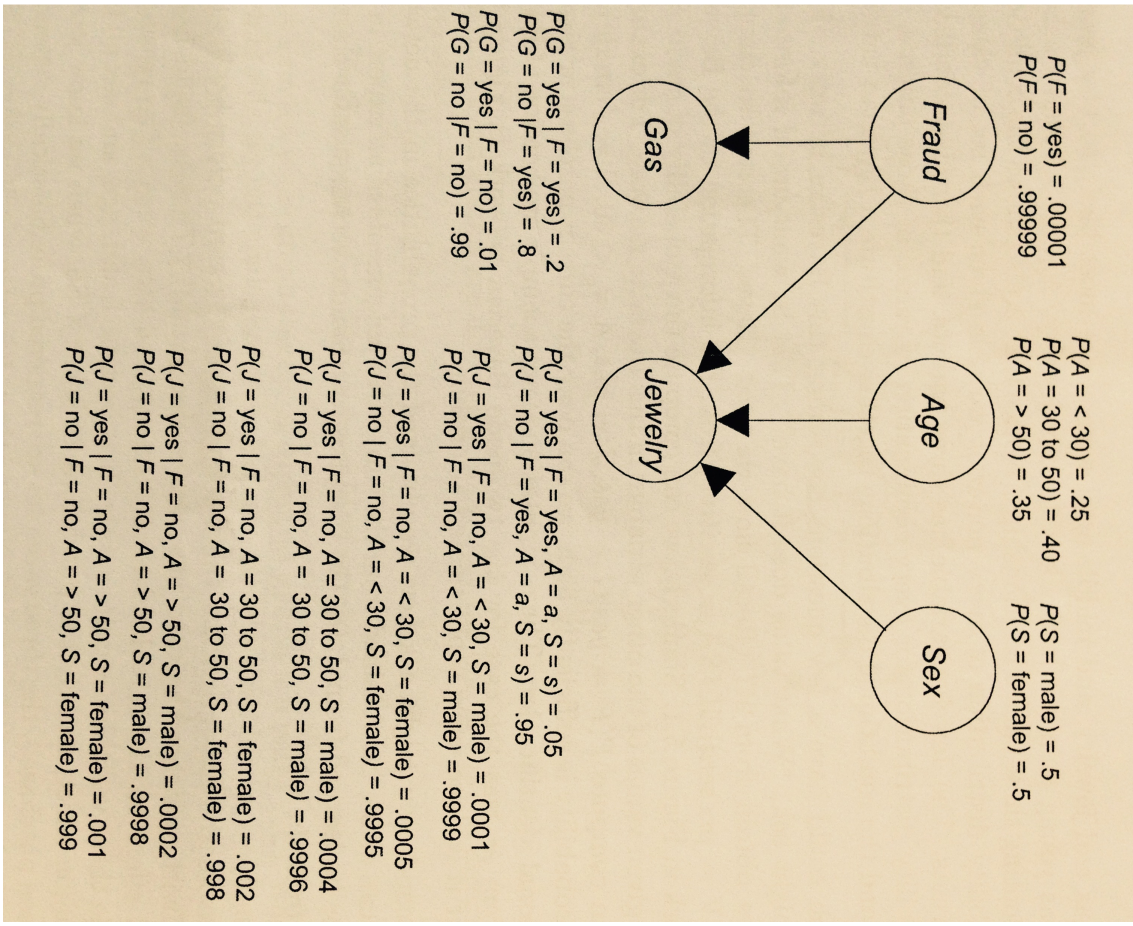


Figura 2: The credit card fraud problem

Use the BN to determine the following conditional probabilities:

(a) $P(F = yes/S = male)$. Is this conditional probability different from $P(F = yes)$? Explain why it is or it is not.

(b) $P(F = yes/J = yes)$. Is this conditional probability different from $P(F = yes)$? Explain why it is or it is not.

(c) $P(F = yes/S = male, J = yes)$. Is this conditional probability different from $P(F = yes/J = yes)$? Explain why it is or it is not.

(d) $P(G = yes/F = yes)$. Is this conditional probability different from $P(G = yes)$? Explain why it is or it is not.

(e) $P(G = yes/J = yes)$. Is this conditional probability different from $P(G = yes)$? Explain why it is or it is not.

(f) $P(G = yes/F = yes, J = yes)$. Is this conditional probability different from $P(G = yes/F = yes)$? Explain why it is or it is not.

(g) $P(G = yes/A \leq 30)$. Is this conditional probability different from $P(G = yes)$? Explain why it is or it is not.

(h) $P(G = yes/A \leq 30, J = yes)$. Is this conditional probability different from $P(G = yes/J = yes)$? Explain why it is or it is not.

5. (**Exercise to deliver individually. Please, attach R scripts**)

Consider Example 3 (see Figure 3) and do the following:

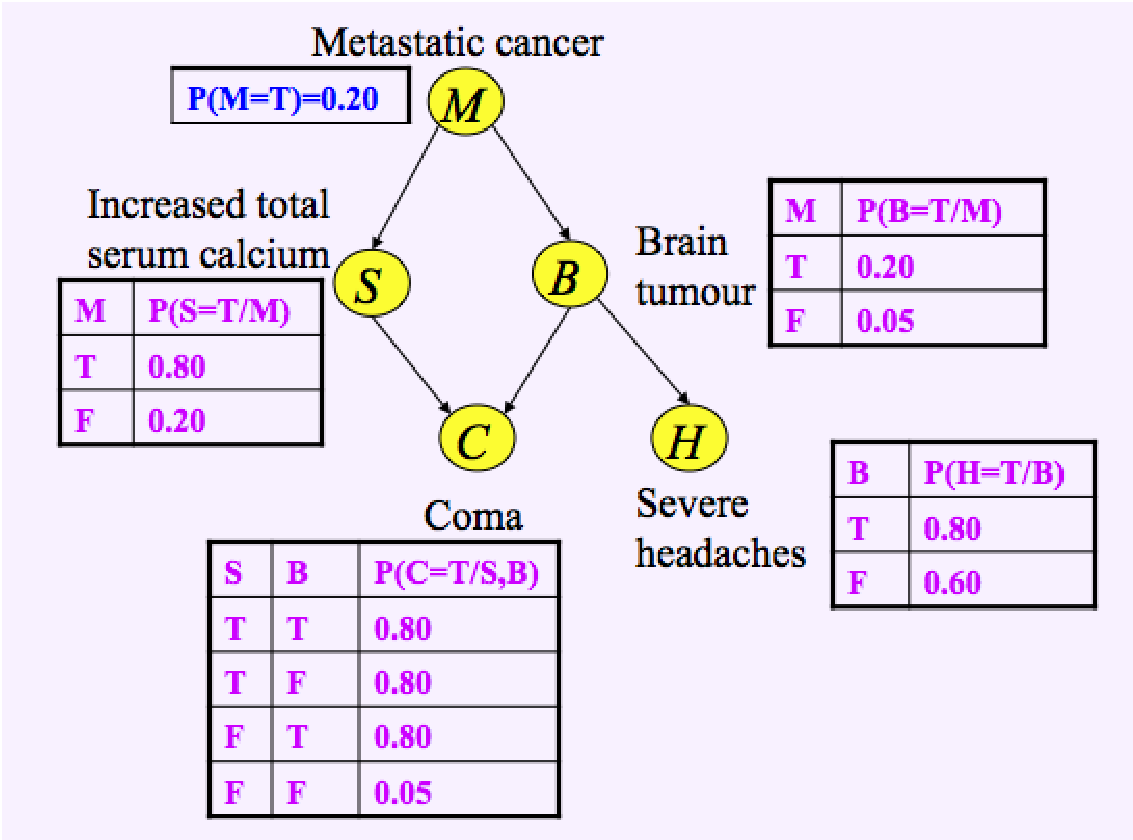


Figura 3: Example 3, Block 2

(a) Using **gRain**, develop the corresponding Bayesian Network and use it to compute $P(M = T/H = F)$ and $P(M = F/H = T)$.

(b) Develop the LS Algorithm to find these two probabilities with R and compare the results with previous item.

(c) Develop the LW Algorithm to find these two probabilities with R and compare the results with previous items.

(d) Compute exactly "by hand" $P(M = T/H = F)$, and compare with (a), (b) and (c).

(e) For the evidence "$H = F$" and the query variable $M$, compute the Kullback-Leibler divergence for the LS Algorithm, and also compute it for the LW Algorithm, and compare them. Which algorithm seems to be better?

6. Joe's x-ray test comes back positive for lung cancer. The test's false negative rate is $f_n = 0.40$ and its false positive rate is $f_p = 0.02$. We also know that the prior probability of having lung cancer in the population is $c = 0.001$ (**prevalence**).

(a) Describe a Bayesian Network and a corresponding query for computing the probability that Joe has lung cancer given his positive x-ray test. What is the value of this probability?

(b) Find a necessary and sufficient condition on the **prevalence** of lung cancer in the population, $c$, that guarantee the probability of cancer to be no less than 10% given a positive x-ray test (if the test's false negative and false positive rates have not been changed).

7. We have three identical and independent temperature sensors that will trigger in:

   - 90% of the cases when the temperature is high,
   - 5% of the cases when the temperature is nominal,
   - 1% of the cases when the temperature is low.

   The probability of high temperature is 20%, nominal temperature is 70%, and low temperature is 10%. Describe Bayesian network and corresponding queries for computing the following:

   (a) Probability that the first sensor will trigger given that the other two sensors have also triggered.

   (b) Probability that the temperature is high given that all three sensors have triggered.

   (c) Probability that the temperature is high given that at least one sensor has triggered.

   Compute the probabilities of (a), (b) and (c) both directly and using **gRain**.

8. Consider the example of the investigator Rich considered in theory (Block 2). Recall that there were three suspects of a murder: David, Dick and Jane, and that the state of belief of the investigator Rich was:

| Killer | Rich P( ) |
|---|---|
| David | 2/3 |
| Dick | 1/6 |
| Jane | 1/6 |

   Suppose now that Rich receives some new evidence that triples his odds of the killer being male.

   (a) What is the new belief of Rich that David is the killer? And Dick and Jane are the killers?

   (b) Recall that Jon is another investigator with the following state of belief, which is different from that held by Rich:

   If Jon were to accept Rich's assessment that the evidence triples the odds of the killer being male, what is the new belief of John that David is the killer? And Dick and Jane?

   (c) We decide to emulate the soft evidence of Rich by a positive sensor reading by using a sensor variable with false negative rate 4%. If the murderer is Jane, with which probability the sensor will give a positive reading?

| Killer | Jon P( ) |
|---|---|
| David | 1/2 |
| Dick | 1/4 |
| Jane | 1/4 |

9. (**Exercise to deliver individually. Please, attach R scripts**)

   We have two sensors that are meant to detect extreme temperature, which occurs 20% of the time. The sensors have identical specifications with a false positive rate of 1% and a false negative rate of 3%. If the power is off (dead battery), the sensors will read negative regardless of the temperature. Suppose now that we have two sensor kits: Kit A where both sensors receive power from the same battery, and Kit B where they receive power from independent batteries. Assuming that each battery has a 0.9 probability of power availability, compute the probability of extreme temperature given each of the following scenarios:

   (a) The two sensors read negative.

   (b) The two sensors read positive.

   (c) One sensor reads positive while the other reads negative.

   Answer the previous questions with respect to each of the two kits. You must use direct computation "by hand" as well as **gRain**.

   In which of the three cases the answer is the same for the two kits? Why?